Jednou nás možná nebude potřeba

Tvář má v komunikaci lidí mimořádně důležitou roli. Jejím prostřednictvím od sebe odlišujeme jednotlivce, z jemných změn v mimice umíme číst emoce a sami je nastavením vlastní tváře vzbuzovat. Část toho je nám vrozena, část se učíme během života. V posledních letech usilují policejní sbory v mnoha zemích světa, abychom tytéž dovednosti, s pomocí technologií a vědy, svěřili strojům – umělé inteligenci. Co nám to přinese?

text **MAREK JANÁČ**

ČESKO PATŘÍ K ZEMÍM, jejichž vědci systémy rozpoznávání obrazu s pomocí umělé inteligence už dlouhodobě studují. Jejich význam totiž s časem poroste nejen v oblasti bezpečnosti, zdůrazňuje profesor brněnského Vysokého učení technického, Martin Drahanský, který u nás patří ke špičce svého oboru.

Když lidé začali zkoumat další živočichy, zejména ty nám nejbližší, občas měli tendenci popisovat jejich inteligenční schopnosti přirovnáním k nějakému vývojovému stádiu člověka. Šimpanzi byli přirovnáváni ke tříletým dětem a podobně. Jak je na tom dnes počítačová umělá inteligence? — Na to se těžko odpovídá, protože umělá inteligence se dnes umí naučit syntetizovat slova a řeč, umí se naučit různé kontexty, znaky, pochopení textu. To v současné chvíli znamená, že v jedné konkrétní úloze může být klidně schopnější než průměrně inteligentní občan kterékoliv země na světě. Umělá inteligence totiž umí třeba přečíst text a zanalyzovat ho tak rychle, že my bychom četli několik let to, co ona zvládne třeba za půl hodiny. V tom už doba určitě pokročila tak daleko, že nástroje dolování dat a nástroje nějakého učení jsou efektivnější než lidský mozek.

Jsme si v něčem podobní? — Snad v procesu učení nesmyslů. Kdyby nám ve škole někdo záměrně pomíchal například jména barev a fialové říkal modrá, zelené červená a podobně, tak se naučíme tuto chybnou interpretaci. Stejné je to s neuronovou sítí. Pokud jí předhodíme nesprávná data, naučí se nesprávně interpretovat skutečnost.

My lidé jsme díky své mozkové kapacitě schopni, pokud jsme trochu zvídaví, čtením a analýzou dalších informací i tento typ klamu odhalit a opravit. Neuronová síť v tom poněkud pokulhává.

Počítačová neuronová síť nezapomíná. — To je pravda, proto musíme věnovat zvýšenou pozornost procesu jejího učení. Pečlivě a často ji kontrolujeme, aby se naučila správně, aby se nepřeučila, nebo naopak nebyla nedoučená. My lidé jsme v tom stále ještě lepší, ale potřebujeme k tomu zase mnoho let a školu.

Když chce člověk využít neuronových sítí, nemusí číst sci-fi, jde na internet či do vlastního chytrého telefonu a třeba zadá neuronové síti Googlu, ať mu ukáže všechny stránky, na kterých je použit obrázek, který ho zajímá. Co dnes rozpoznávání obrazů vlastně umí? — Odpověď není opět jednoduchá. V podstatě vždycky musíte dát neuronové síti vzory, které se naučí. Pokud hledáte to, co už někde v podobě snímku (třeba květináče s kvítkem) existuje, v takovém případě je šance na nalezení stejného snímku poměrně vysoká. Pokud ale stejné kvítko vyfotíte jen z trochu jiné pozice, nebo tam budete mít jiné světelné podmínky, šance na nalezení dramaticky klesne.

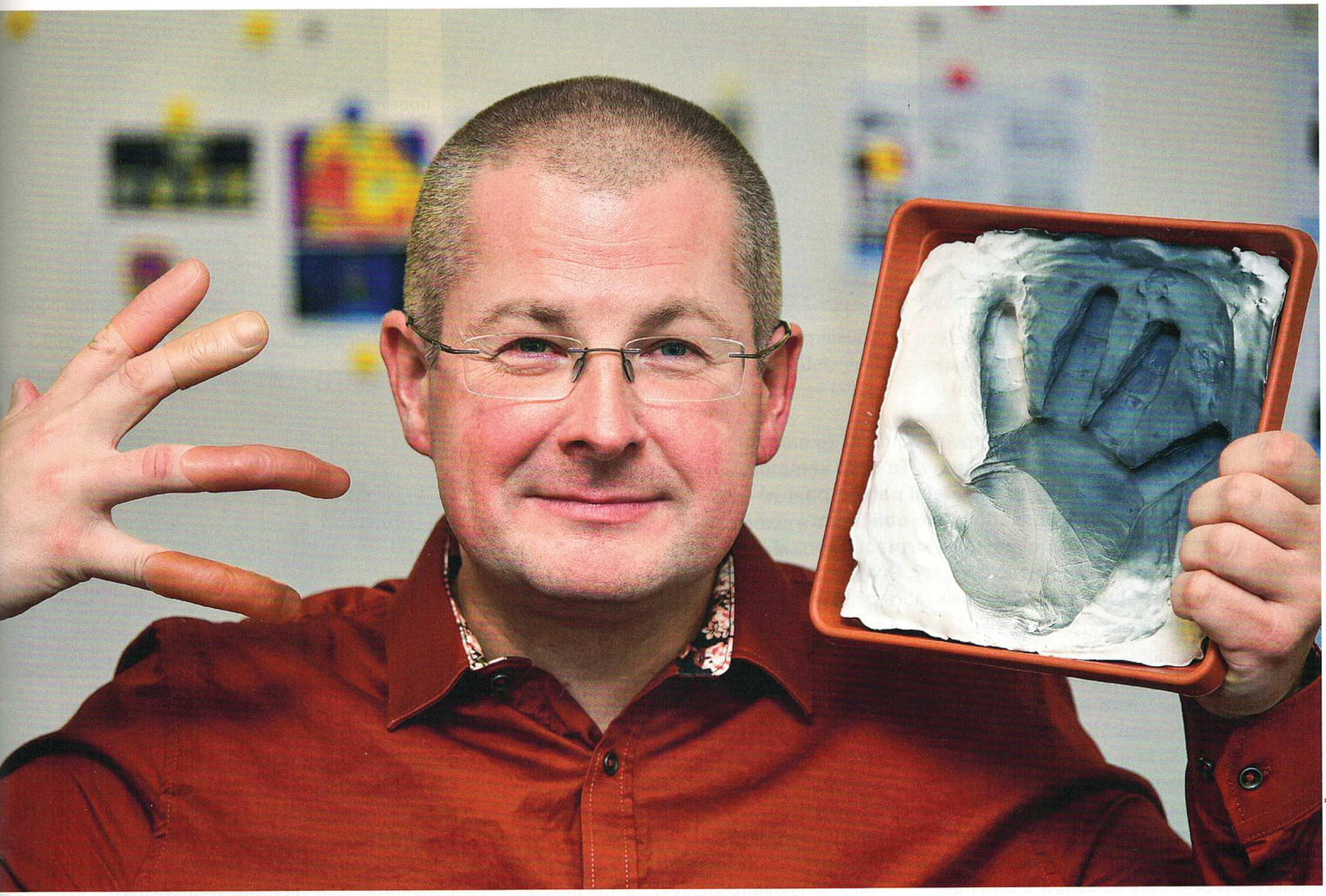
Jinými slovy matematický překlad té jedné fotografie je identický ať ji zvětšíte či zmenšíte, zatímco jakmile ji nasnímáte z jiného úhlu, bude to jiný matematický obraz. — Je to tak. Ten přepočet do digitální formy, extrakce příznaků, které tam jsou, tak už to vypadá jinak. Když budete hledat nějaké děvče v plavkách, které jste předtím viděl na pláži a vyfotografoval si ji, a nahrajete ji do Googlu, šance, že vám najde přesně totéž děvče je téměř nulová. Ukáže vám tisíce fotografií jiných děvčat na pláži, ale pravděpodobně ne tu, kterou jste hledal.

Tím se dostáváme k hledání obličeje. — Jednou z dnes asi nejlepších aplikací na hledání obličeje je algoritmus společnosti NEC. Vyhrál dokonce i testování pořádané americkým institutem NIST (US National Institute for Standards and Technology),[1] takže patří ke špičce v oboru. Funguje perfektně, umí výborně rozpoznávat obličeje, zejména čelně nasnímané. Problém začíná v okamžiku, kdy už mají nasnímaní lidé trochu nakloněnou či otočenou hlavu. Dochází k nepřesnostem a rozpoznávání vypadává. Fungovalo by to ještě tak při hledání snímku Brada Pitta. Šance, že na prvních 100 místech bychom měli správné výsledky by byla poměrně veliká, ale to jenom díky tomu že byl ve filmech nasnímán ze všech možných úhlů, proto je šance na nalezení shody poměrně vysoká. V okamžiku, kdy bychom použili mou profilovou fotografii, je docela možné, že ji systém také nalezne, ale mnohem pravděpodobněji bych čekal jako výsledek docela jiného člověka, který má podobně šišatou hlavu, jako já, má brýle a rašící vousy. Takže tam je otázka, co použít.

Jeden z nápadů, které má vaše skupina, je využití skenu 3D obličeje. — Ano, hlava se natočí do správné pozice, a výsledek se porovná s dvourozměrným snímkem, tím můžeme daného člověka rozpoznat. Stoprocentní spolehlivost je ale ještě práce na dlouhou dobu. Potíže zatím dělá zejména rozpoznání trojrozměrné pozice hlavy na dvourozměrném snímku, to ještě opravdu spolehlivé není. Co dnešní algoritmy dokážou, je kupříkladu rozpoznání obličeje ovlivněného stárnutím. To znamená, že když budu hledat podle nějaké starší fotografie, šance že mě algoritmus stále najde je poměrně veliká.

Ale je to zase specializovaný typ algoritmu, který je určený na hledání obličeje. Co když potřebuji zjistit, jestli některý

1) Programy, účastnící se testu, dostávají ke zpracování velmi rozsáhlé databáze, obsahující fotografie více než 1,8 milionu osob.



Snímek Igor Šefr, VUT v Brně

Prof. Ing. Dipl.-Ing. MARTIN DRAHANSKÝ, Ph.D., (*1978) vystudoval FEI VUT v Brně a FE FernUniversität v Hagenu (2001), získal Ph.D. (2005), habilitaci (2009) a profesorský dekret (2017) na FIT VUT v Brně. Biometrickými systémy se zabývá od roku 1999.

z lidí na obrázku nemá v ruce zbraň? — Na to jsou jiné specializované algoritmy. Mohou, ale kupodivu také nemusejí, využívat neuronové sítě. Dokonce je dost často ani nevyužívají, ale snaží se ze snímků vytáhnout nějaké zajímavé příznaky tak, aby potom bylo možno ten daný artefakt, který hledáte, rozpoznat a říci, jestli na snímku skutečně je a kde je. Takové algoritmy už jsou, jde jenom o to, abychom je dotáhli do vysoké spolehlivosti. Když máme relativně kvalitní snímky, spolehlivost může být mezi 80–90 %, s poklesem kvality zdrojových snímků však tyto algoritmy velmi chybují a bez lidské asistence se neobejdou.

Co brání spolehlivost zvýšit? — Ohromné množství kvalitních trénovacích dat. Abychom na nich neuronovou síť natrénovali, potřebujeme jí říci: „Na tomto místě je vidět, jak člověk tasí zbraň v davu, nauč se to, a pokud někde v davu uvidíš podobné lidské chování, je to podezřelé, a dej nám znamení, že se něco děje."

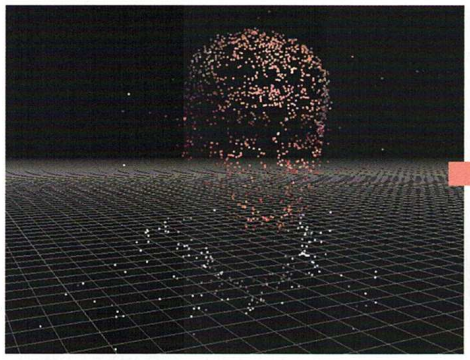
Nevím sice o tom, že by zatím takový konkrétní algoritmus existoval, ale řada jiných už existuje. Nejsou sice ještě úplně spolehlivé, ale už se blížíme k opravdové použitelnosti v praxi, kde nám dávají v drtivé většině případů správné výsledky.

Tam je asi problém, že zbraně na ulici lidé v Evropě nevytahují tak často. — Naštěstí máme filmové databáze, kde je lidí tasících zbraň, poměrně dost. Bohužel u filmových databází je velká nevýhoda, že jde ve většině případů o hrané scény a že část z nich jsou i westerny, kde se pistole nosí viditelně a tasí se při každé příležitosti. Spadá nám do toho i Matrix a podobná díla, kde se zbraně také tasí za úplně jiných podmínek.

U kterých žánrů bychom měli neuronové síti zakázat sledování a které povolit? — Zakázal bych už zmíněné westerny. Povoleny by měly být kriminálky, kde to už je aspoň trochu realistické. Ale to zase musí udělat člověk. Neuronová síť sama o sobě neví, jak má poznat, kdy je tasení zbraně westernové, kdy matrixové, a kdy něco opravdu reálného. Data je třeba anotovat člověkem.

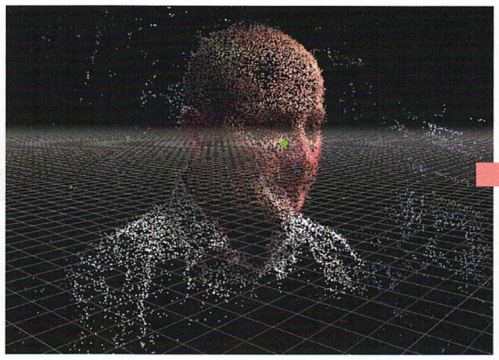
Něco jiného také bude, když člověk tasí pistoli, bazuku, nůž, případně když navíc kulhá, či jde o ženu. Každý ten problém se patrně musí řešit samostatně. — V zásadě - pokud chcete na zakázku vyrobit software tohoto druhu, musíte říct, na jakých datech to má fungovat. To znamená, že pokud si zvolíte jako cíl rozpoznat tasení zbraně za jakýchkoliv podmínek, tak já řeknu, že to udělat nelze. Důvody jste sám popsal. Doplnil bych ještě problém oblečení. Pokud mám na sobě dlouhý kabát až na zem, je to něco jiného, než mám-li na sobě tílko a trenýrky a mám zbraň schovanou někde za gumou, za zády. Variabilita zrovna u tasení zbraně je tak obrovská, že je v tuto chvíli téměř nemožné natrénovat neuronovou síť tak, aby situaci správně rozpoznala. Nemáme dostatek dat.

To znamená, že protizbraní proti neuronovým sítím a umělé inteligenci je kreativita. — Určitě. Pokud budete kreativní a vytvoříte něco, co ta neuronová síť neumí, tak vás buď vůbec nenajde, anebo najde něco docela jiného, než by měla najít. Kupříkladu když budete na ulici a nasadíte si třeba hlavu jelena s nějakým nápisem mezi parohy, tak neuronová síť bude zmatena a řekne si, že tam není žádný obličej, bude považovat obrázek za snímek zvířete a zahodí ho. Ona neví, že se může stát, že se někde prochází člověk, který má na hlavě nasazenou jelení hlavu. Na druhou stranu, pokud bychom vzali pachatele, který si nasadí na hlavu masku nějakého zvířete, tak může být velmi pravděpodobně ještě předtím zabrán nějakou jinou kamerou, kde si ji ještě nenasadil, takže sice potom vidíme, že se někde

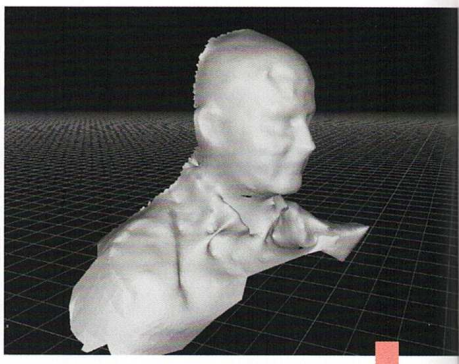


Snímky Tomáš Goldmann

řídké pole bodů



husté pole bodů



3D model

3D model s texturou

NEJVĚTŠÍ PROBLÉM při rozpoznávání obličejů ze snímků kamer je orientace hlavy snímané osoby. Ta bývá většinou v jiné pozici, než jak ji databáze zná (čelní pohled pasové fotografie). Tým prof. Drahanského proto obličejům vytváří trojrozměrné hlavy z naskenovaných informací, které si virtuální realita natáčí a porovnává shody. Pro každou rozpoznávanou osobu si pak virtuální realita připravuje samostatný model.

v davu objevil jelen, jsme schopni stopovat jelenův pohyb a v nějaké jiné kameře spatřit obličej člověka, který si masku nasadil. Což je zase výhoda kamerových systémů. Pokud hledáme pachatele, tak ho takto můžeme dosledovat, ale musíme mít co největší pokrytí. Pokud máme ve městě pouze jednu kameru, která se dívá na banku, a tam nám někde z boku přicupitá člověk s hlavou jelena, tak už s tím nic nenaděláme. Proto prostě potřebujeme co nejvíc dat, abychom měli co prohledávat.

V listopadu loňského roku požádala česká policie pražský magistrát o povolení aktivovat na šesti kamerách v centru hlavního města automatické rozpoznávání obličejů. Z tohoto pohledu se dá žádost brát i jako pokus o trénování neuronové sítě na reálných datech. — V zásadě ano. Tady je třeba dodat, že většina těch dat se asi bude zahazovat. Jde totiž o to, aby vznikl algoritmus, vyhledávající nějaké - z bezpečnostního hlediska problematické - chování. Ten algoritmus by si to už mohl spojit s libovolnou osobou, která by se hledaným způsobem chovala. On by z těch dat vytáhl obrovské množství informací, zejména o stylu chování, a mohl by se to naučit.

To by ale mohlo vést ke vzniku nějakého hypotetického PreCrimu, jako ve filmu Minority Report. Policie bude schopna předvídat chování lidí? Bude možné, aby sledování přineslo tento typ informací, na jejichž základě bude možno předpovědět chování nějakého člověka? — Já s tím docela počítám. Dokonce bych i chtěl, aby to vedlo k poznání: „Ano, tyto osoby se chovají podezřele, je velmi pravděpodobné, že plánuji útok, vyhlíží si oběť nebo něco takového." V tom případě by se tam někde měla objevit policejní hlídka. Stačilo by jim pak zkontrolovat doklady. Domnívám se, že pokud by se jim to za noc stalo třikrát, tak si nějaký útok rozmyslí, čehož bychom měli chtít docílit. Určitě nechci prekriminalizovat společnost, aby se nestalo, že pokud někdo půjde z diskotéky domů, bude opilý a bude se chovat podezřele, aby ho někdo okamžitě sebral a odsoudil na 15 let do vězení. To je samozřejmě nesmysl. Jde ale o to, aby se těm, kteří páchají trestnou činnost, tato zavrženíhodná aktivita co nejvíce znepříjemnila. Časem zjistí, že vůbec nemá cenu takovou trestnou činnost páchat a bude to vyřešeno. Myslím, že k tomuto typu prevence by to mohlo směřovat.

Řada lidí má ale právě z takového sledování svých kroků obavu. — Chápu lidi, kteří mají strach z toho, že mohou být nafilmováni na nějakém konkrétním místě a kvůli tomu se bude vědět, kudy se pohybovali. Jenomže ti lidé si neuvědomují, že pokud vlastní a používají mobilní telefon, už nyní dávají operátorovi možnost naprosto přesně sledovat svou pozici kdekoliv a kdykoliv. Je to většinou velmi osobní, protože existuje podepsaná smlouva s operátorem, na které je číslo telefonu, adresa majitele, číslo bankovního konta, datum narození. Policie nebo další složky mohou v případě oprávněného zájmu naprosto přesně zjistit, kde se ten člověk nacházel a co dělal.

Mně osobně strach opustil, protože si uvědomuji, co všechno je dnes schopná umělá inteligence najít a vycucat za data. Myslím si, že už se tomu neubráníme,

simulace kožních chorob (ekzému, bradavic, lupénky)

simulované podvrhy otisků prstů (změny tvaru, vzduchové bubliny)

Obrázek Ondřej Kanich

ROZPOZNÁVÁNÍ OBRAZU má praktický dopad na bezpečnost systémů, chráněných otisky prstů. Pokud útočník použije k ošálení systému podvrh otisků prstů (dole), obsahuje obraz buď změny tvaru, anebo vzduchové bublinky. Běžné poruchy kůže, vzniklé nejrůznějšími nemocemi (nahoře), vypadají jinak.

protože cokoliv, co děláme - a je to alespoň trošku provázáno elektronicky a někde se něco ukládá -, tak tím sdělujeme umělé inteligenci o sobě naprosto vše a myslím si, že se tomu nejsme schopni bránit. Prostě tak to je, společnost taková chce být, a pokud nechceme zastavit používání mobilních telefonů a veškeré elektroniky, kterou máme, zejména internet věcí, jako jsou nejrůznější inteligentní hodinky a podobně, pak se s tím musíme smířit.

Máte představu, kdy přijde něco takového v podobě systémů ochraňujících lépe bezpečnost v naší společnosti? — Možná to bude ještě dva, tři roky trvat, ale myslím, že jako první budou nasazeny kamerové systémy s rozpoznáváním obrazu a automatickými alarmy třeba na letištích a v dalších veřejně přístupných místech. Už nyní je tendence využít kamerové systémy nejenom k tomu, že najdou nějakého člověka podle obličeje, ale už se řeší opuštěné zavazadlo, které by mohlo být potenciálně výbušninou. To už funguje. Domnívám se, že dohledání nějakých osob z různých pohledů, tedy i bočních záběrů na obličej, to věda časem také přinese, ať už od nás, nebo konkurenčních týmů. Podle mne jde o horizont dvou až tří let, kdy budou kamerové systémy tuto schopnost mít.

Bavíme se sice o bezpečnosti, ale medicína je také obor, kde by tyto technologie našly uplatnění. Sami pracujete na systému automatického vyšetřování oka pro lékařské účely. Nenastane to tam o něco dříve? — Pevně doufám, že to bude už letos. Problém je s různorodostí patologických projevů na sítnici oka. Je obrovská, podobně jako při rozpoznávání tasení zbraně. Naučit správně všechny algoritmy, aby byly schopny detekovat korektně patologické nálezy v sítnici oka, je díky tomu mimořádně náročná úloha.

Časem tedy přijdu k lékaři, ten bude mít vaši kameru, která sama najde mé oko, zaměří se na něj, naskenuje oční pozadí, a lékaři během chvilky řekne, že má podezření na rozvíjející se rozpad sítnice, nebo něco takového? — Takto pojaté zařízení by bylo zbytečně drahé. U praktického lékaře postačí obyčejná fundus kamera, vybavená naším algoritmem, který bude schopen říci: ano na sítnici oka vidím některé typy změn, které tam asi být nemají, třeba tvorbu nových arteriálních větvení a podobně. V takovém případě obvodní lékař zpozorní. Nebude to sice umět vyhodnotit, ale pošle pacienta za očním specialistou, který potom stanoví diagnózu.

Asi nejčastějším onemocněním, jak vyplývá ze zpětné vazby od lékařů, je cukrovka a věkem podmíněná makulární degenerace. Takže ta šance, že se dá nemoc objevit ještě předtím, než se vůbec něco s okem stane, dnes záleží na tom, jak moc chodíte k lékaři nechávat si udělat rozbory krve. Pokud byste si však při nějaké obecné prohlídce nechal jednoduše proměřit oko touto kamerou, dá se poměrně rychle zjistit v jakém stavu vaše sítnice je, co se tam děje, a obvodní lékař problém zachytí včas. Náš software by jednou měl přibližně stanovit i obsah problému, ale to není tak důležité. Finální rozhodnutí vždy dělá odborný lékař - oftalmolog.

Je složité naučit kameru, respektive její software, vyhodnotit oko? — Já bych řekl, že to je skoro srovnatelné, protože pro nás je zásadní problém zase mít hromadu dat, nejlépe anotovaných lékaři, abychom věděli, jak vypadá drúza, jak krvácení, jak rozšíření drobných cévek oka. Nestačí nám jeden snímek, potřebujeme ty případy ve všech podobách, v jakých se v oku mohou vyskytnout. Naštěstí u sítnice máme poměrně rozsáhlé databáze, některé z nich dokonce i popsané lékaři, ale bohužel ne všechny, což je škoda.

Mají se očaři bát o práci? — Dnes bych o to neměl strach. Než se nějaký stroj naučí rozpoznávat a případně operovat oko, uběhne ještě hodně dlouhá doba. Udělat kameru, která vám vyfotí oko, stanoví diagnózu, rozhodne o léčbě a hned vám ji i provede třeba laserem, toho bych se zatím bál. Sám, jako člověk, který to tady s kolegy vytváří, bych do toho nešel, protože to není odzkoušené. Myslím, že potřebujeme nasbírat obrovské množství dat a otestovat to několik let v praxi, abychom viděli, že to opravdu spolehlivě funguje. V následujících nejméně pěti až deseti letech to určitě nenastane. Pokud se ale bavíme o dvaceti a více letech, pak jsem si jist, že k tomu robotizace určitě dospěje.

S tím, jak se technologie rozpoznávání obrazu budou zlevňovat, budou dostupnější i mimo zdravotnictví a bezpečnostní složky. Vy už dnes v Přírodovědeckém muzeu Národního muzea pracujete na systému, který by měl pomoci vědcům při prezentaci jejich poznatků na výstavách. Co to může přinést? — Ten potenciál je poměrně velký. Požadavek je zjistit, co návštěvníky nejvíc zajímá.

To se neví? — Muzejníci mají několik hlavních možností, jak návštěvníka zaujmout. Od vystavení unikátních sbírkových předmětů přes zařazení interaktivních prvků, s nimiž si lidé mohou hrát a mohou si na ně nějakým způsobem sáhnout, po infopanely, u nichž lidé stojí a čtou si v nich, nebo sledují nějaký text či audiovizuální rekonstrukce. Například podoby pravěkého života. A teď jak zjistíte, který z těchto prvků návštěvníky zajímá nejvíce?

Zjistíme pozorováním. — Jistě, ale abychom našli nějakou exaktnější odpověď, pošleme tam dobrovolníky, kteří si budou dělat čárky podle toho, kolik lidí kde stojí, čte si něco a podobně. My chceme jít jinou cestou - použít kamerové systémy a automatické vyhodnocení. Chceme zjistit, kde se pohybují návštěvníci, kde se zastavují, a v budoucnu i případně na co se dívají, co čtou, jestli čtou text anglický nebo český, případně jestli ho vůbec čtou. Pak můžeme konečně muzejníkům říci, co je pro lidi skutečně nejlepší.

Anebo můžete zjistit, že k nějaké konkrétní vitríně lidé chodí málo. Tak ji proměníte, dáte tam něco, něco jinak uděláte a zjišťujete jestli k ní chodí víc lidí, případně kam se dívají a co je zajímá. Zda se úpravou expozice proměnila tak, aby zajímala víc lidí. — Přesně tak. A pokud se zjistí, že text nikoho nezaujal, je zapotřebí analyzovat proč. Jestli byl moc malý, nebo moc dlouhý a lidé ho nechtěli číst. To asi nezjistíme z toho videa, ale minimálně se bude vědět, že je něco špatně, protože lidé texty nečtou.

Čím je to nové v muzejnictví? — Pokud vím, tak zatím neexistuje nikde na světě systém, který by byl založen na kamerách. Zatím je mají jako dohledové zařízení a zkoušeli zjišťovat chování návštěvníků pomocí různých technologií, jako jsou technologie UWB, nebo bluetooth, brýle pro sledování pohybu očí a podobně. Každý z těch systémů má nějakou zásadní nevýhodu. Jeden je drahý, druhý nepřesný, třetí návštěvníky obtěžuje a nechtějí ho používat.

Námi navrhovaná technologie je pasivní systém, od návštěvníků nic nevyžaduje, vypadá jako dohledový systém, který tam stejně je, proto nikoho neruší a data se nikde neukládají. Lidé tedy nemusejí mít strach, že by byli někde natočení a někde se to mohlo zobrazit. V okamžiku, kdy se data vyhodnotí, se jednoduše smažou a už nejsou. Vůbec je neuchováváme, díky čemuž i návštěvníci mohou být v klidu, že se vlastně vůbec nic neděje, protože máme jenom nějakou statistiku a žádná osobní data uložena nejsou.

Až jednou neuronové sítě a umělou inteligenci naučíme všechno, co známe, co nás čeká? — Pokud toho bude někdy v budoucnu schopna, bude umět zjistit i nějaká abstraktní data, která my lidé, ještě nevidíme. Pak už nebudeme potřeba. Budeme maximálně vyrábět elektřinu na šlapacích kolech, aby mohla umělá neuronová síť fungovat.